

# Reliably Detectable Flaw Size for NDE Methods that Use Calibration

Ajay M. Koshti, D. Sc. PE  
NASA Johnson Space Center, Houston

March 25-29, 2017

**SPIE Smart Structures/NDE 2017**  
**Portland, OR, U. S. A.**

- Acknowledgements
- References
- Introduction
- Background, Two Types of POD Datasets
- Three Cases of POD Datasets and POD Objectives
- Case 1: Cracks in Flat Plates (One Dataset)
- Case 2: Crack and Notch Response Data Sets on Flat Plates (Two Datasets)
  - Method A and B
- Case 3: Detection of Cracks in Complex Geometry Parts (Three Datasets)
  - Method A and B
- Conclusions

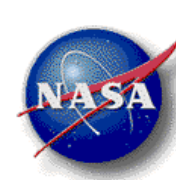


# Acknowledgements



- Buzz Wincheski (NASA, Langley Research Center),
  - David Stanley (NASA Johnson Space Center),
  - Charles Nichols (NASA White Sands Test Facility),
  - James Walker (NASA Marshall Space Flight Center),
  - Michael Suits (NASA Marshall Space Flight Center) and
  - Craig Bryson (NASA Marshall Space Flight Center)
- provided valuable review comments to this work.

- [1] “Nondestructive Evaluation System Reliability Assessment,” MIL-HDBK-1823A, Department of Defense, USA, (2009).
- [2] Annis, Charles, Statistical Engineering, “mh1823 POD Software V5.2.1,”  
<http://www.statisticalengineering.com/mh1823/>.
- [3] “Standard Practice for Probability of Detection for Hit/Miss Data,” ASTM E 2862-12, ASTM International.



# Introduction



- Probability of detection (POD) analysis is used in assessing reliably detectable flaw size in nondestructive evaluation (NDE).
- MIL-HDBK-1823<sup>1</sup> and associated mh1823<sup>2</sup> POD software gives most common methods of POD analysis.
- In this paper, POD analysis is applied to an NDE method, such as eddy current testing, where calibration is used.
- NDE calibration standards have known size artificial flaws such as electro-discharge machined (EDM) notches and flat bottom hole (FBH) reflectors which are used to set instrument sensitivity for detection of real flaws.
- Real flaws such as cracks and crack-like flaws are desired to be detected using these NDE methods.
- A reliably detectable crack size is required for safe life analysis of fracture critical parts.
- Therefore, it is important to correlate signal responses from real flaws with signal responses from artificial flaws used in calibration process to determine reliably detectable flaw size.



# Background, Two Types of POD Datasets



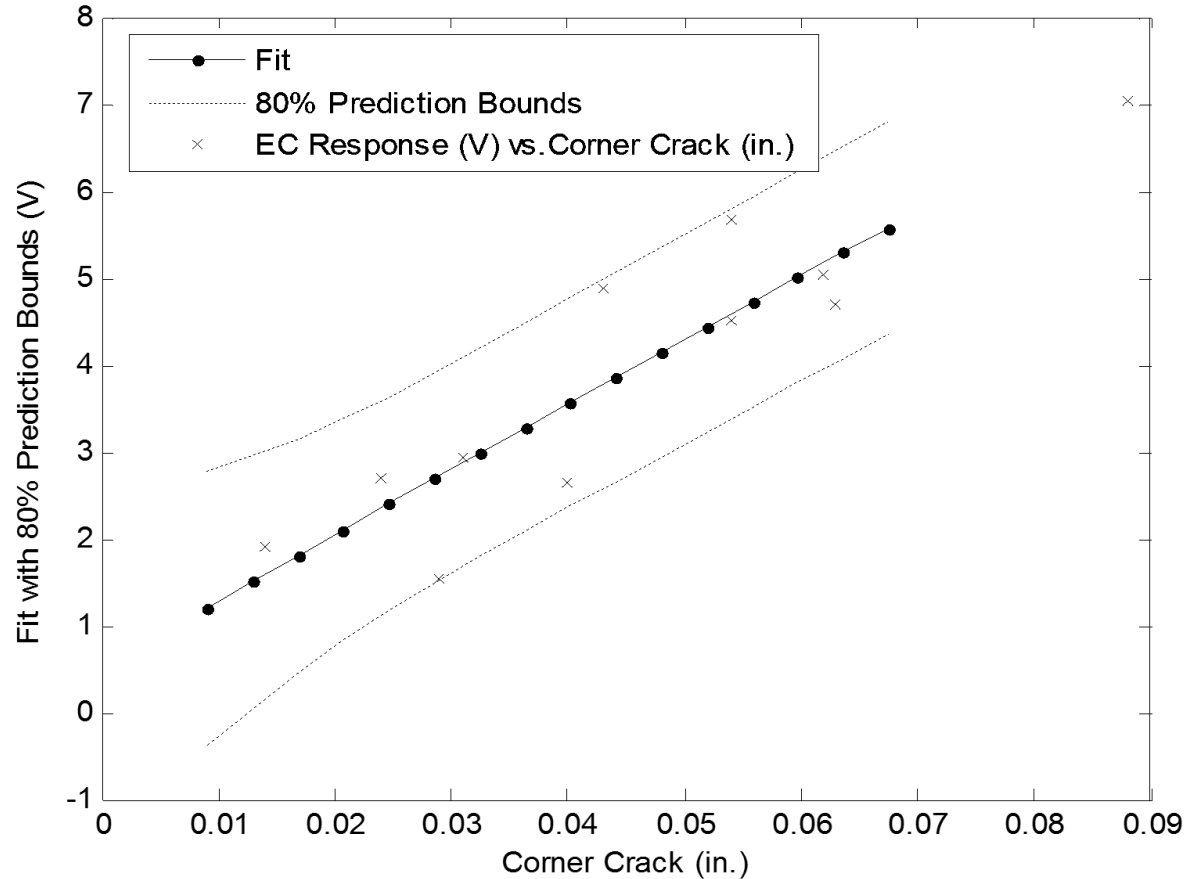
- MIL-HDBK-1823<sup>1</sup> and associated mh1823<sup>2</sup> software cover two types of datasets.
- First type of dataset is signal response  $\hat{a}$  (read as a-hat) versus flaw size “ $a$ ”.
  - The  $\hat{a}$  (y-axis) versus “ $a$ ” (x-axis) data may be transformed using logarithm function along appropriate axes, if needed, to create linear correlation around the decision threshold,  $\hat{a}_{decision}$ .
  - A generalized linear model (GLM) is fitted to the transformed data for analysis.
  - Here, noise data is taken separately to define noise distribution.
  - Noise is same as signal response from part where there is no flaw.
  - Noise data is used to determine false call rate or probability of false calls (POF).
- Second type of dataset is called hit-miss data, which contains flaw size and corresponding detection result i.e. hit or miss.
  - Hit has numerical value of 1 and miss has numerical value of 0.
  - Here, false call data is noted to determine false call rate using Clopper-Pearson binomial distribution function.
  - Normally, POD increases with flaw size and POF decreases with flaw size.
  - POF value shall be within certain limit to prevent adverse impact on cost and schedule.
  - ASTM E 2862<sup>3</sup> also provides the hit-miss POD data analysis method that is consistent with MIL-HDBK-1823.



# Three Cases of POD Datasets and POD Objectives



- Case 1 has a single dataset of signal responses on cracks (dataset 1).
  - The data is used to set eddy current decision threshold.
- Case 2 uses two signal response datasets.
  - One dataset has signal responses on cracks in flat plates.
  - The second dataset (dataset 2) has signal responses from EDM notches in flat plates of the same material.
  - Case 2 assumes that part surface contour has no effect on EC flaw response and flat plates are same as the part for inspection purposes.
- Case 3 assumes that cracks are to be detected in complex geometry part such as an orbital tube weld (OTW).
  - It also assumes that known size cracks are too expensive to fabricate in samples with part configuration.
  - Therefore, a set with real cracks in part configuration, although most desired, is not available.
  - Instead, a configuration (e.g. OTW) specimen set (dataset 3) with EDM notches, simulating cracks, is available.
  - Also, a set with notches in flat plates of part material (dataset 2) and a set with cracks in flat plates of part material (dataset 1) are available. Thus, here we have three signal response datasets, one each from cracks in flat plates, notches in flat plates and notches in part configuration specimens.
- POD Objectives
  - POD analysis takes into account variance in data
  - Method A: Calculate reliably detectable flaw size for selected decision threshold, and
  - Method B: Calculate decision threshold for desired reliably detectable crack size.



- From above plot, half interval of the prediction bounds,  $\Delta a_{p90}$  can be directly computed as half of range between the two bounds at a given decision threshold as follows,

$$\Delta a_{p90} = (a_{90L} - a_{90U})/2 \quad (1)$$

- Lower and upper 95% confidence bounds at a given decision threshold are given by,

$$a_{90L/95} = a_{90L} + \Delta a_{c90L/95}, \quad (2)$$

$$a_{90U/95} = a_{90U} - \Delta a_{c90U/95}. \quad (3)$$

See next chart

Fig. 1: Flat plate crack response and prediction bounds



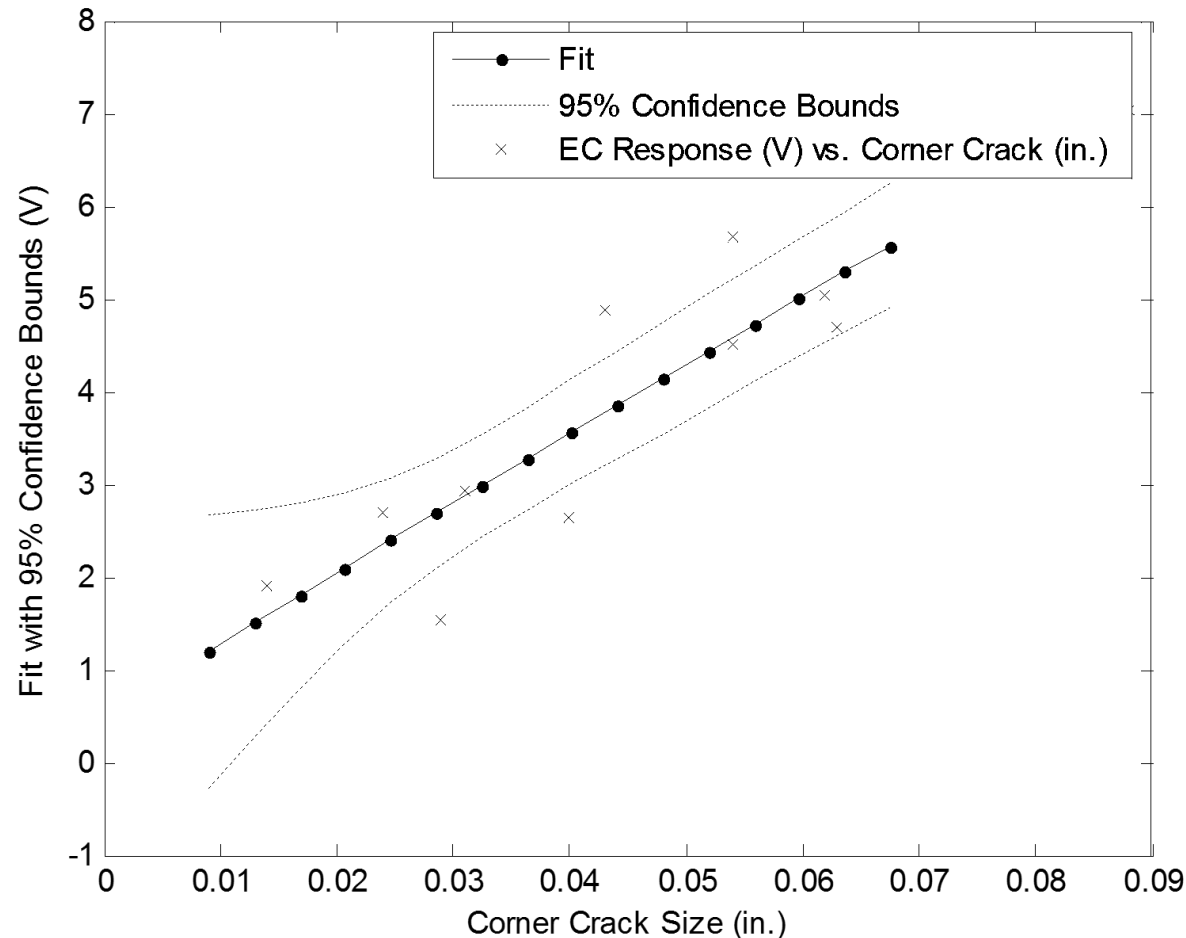


Fig. 2: Flat plate crack response and confidence bounds

- $\Delta a_{c\ 90L/95}$  and  $\Delta a_{c\ 90U/95}$  = half of 95% confidence interval at given flaw sizes  $a_{90L}$  and  $a_{90U}$  respectively.
- Subscript “c” is used to indicate confidence interval.
- From above plot,  $\Delta a_{c\ 50/95}$  can be directly computed as half of range between the two bounds at a given decision threshold as follows,

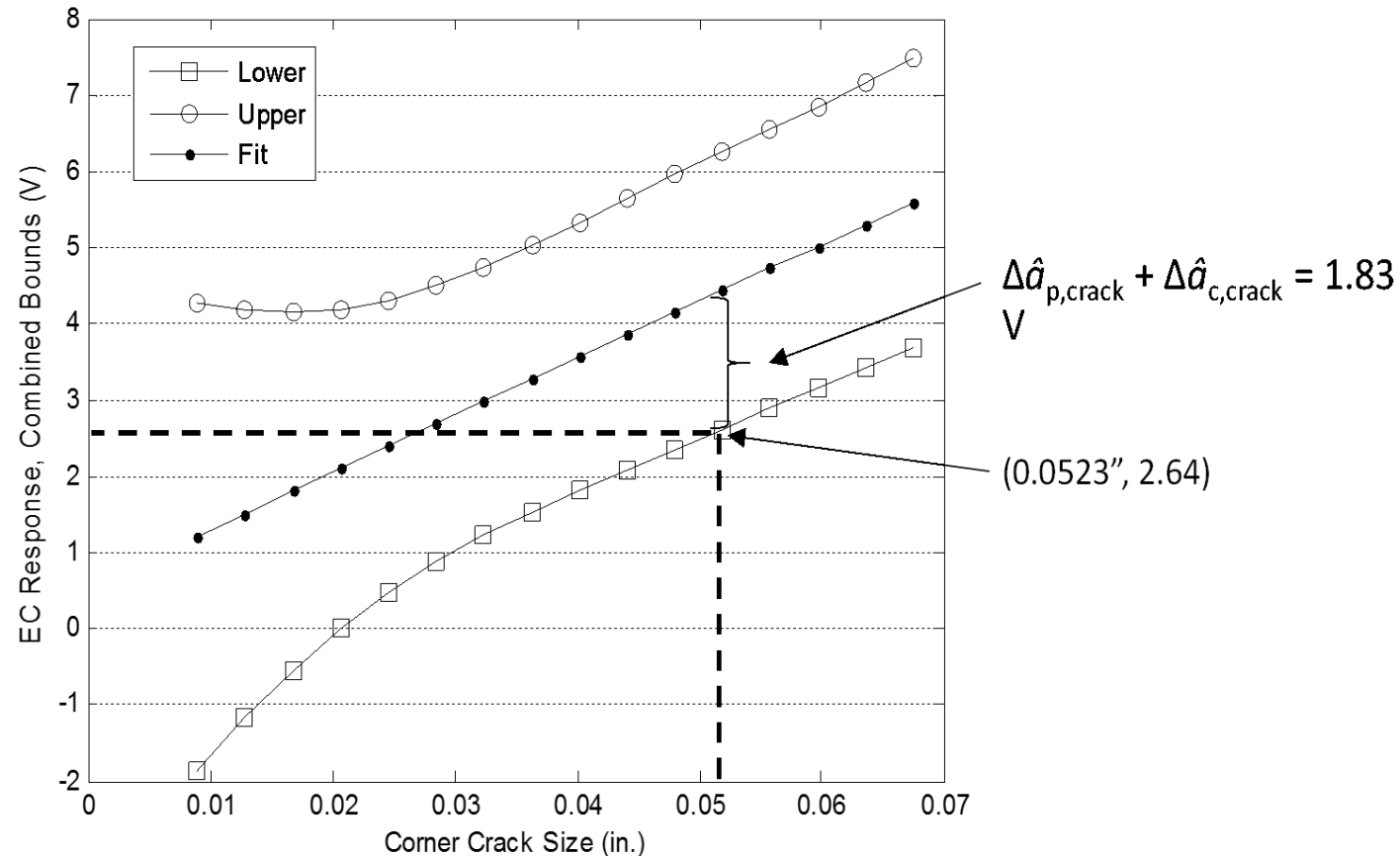
$$\Delta a_{c\ 50/95} = (a_{50/95L} - a_{50/95U})/2. \quad (4)$$

- If slopes of prediction bounds  $a_{90}$  and corresponding confidence bounds  $a_{50/95}$  do not differ, indicating that variance in data and confidence interval do not change significantly with small change in decision threshold, then following approximation can be justified.

$$\Delta a_{c\ 90L/95} = \Delta a_{c\ 90U/95} \approx \Delta a_{c\ 50/95}. \quad (5)$$



# Case 1: Cracks in Flat Plates, Combined Bounds on Curve Fit



- The difference would be noticeable when slope of upper and lower  $\hat{a}_{90}$  versus  $a$  curve differ from corresponding  $\hat{a}_{50/95}$  versus  $a$ .

- Therefore, substituting  $\Delta a_{c\ 90L/95}$  and  $\Delta a_{c\ 90U/95}$  by  $\Delta a_{c\ 50/95}$  we get,

$$a_{90/95L} \cong a_{90L} + \Delta a_{c\ 50/95}, \quad (6)$$

$$a_{90/95U} \cong a_{90U} - \Delta a_{c\ 50/95}, \quad (7)$$

$$a_{90L/95approx} \cong a_{90L} + \Delta a_{c\ 50/95}, \text{ and } (8)$$

$$a_{90U/95approx} \cong a_{90U} - \Delta a_{c\ 50/95}. \quad (9)$$

Fig. 3: Flat plate crack response fit with combined bounds



# Case 1: Cracks in Flat Plates, Calculation Using mh1823 POD Software

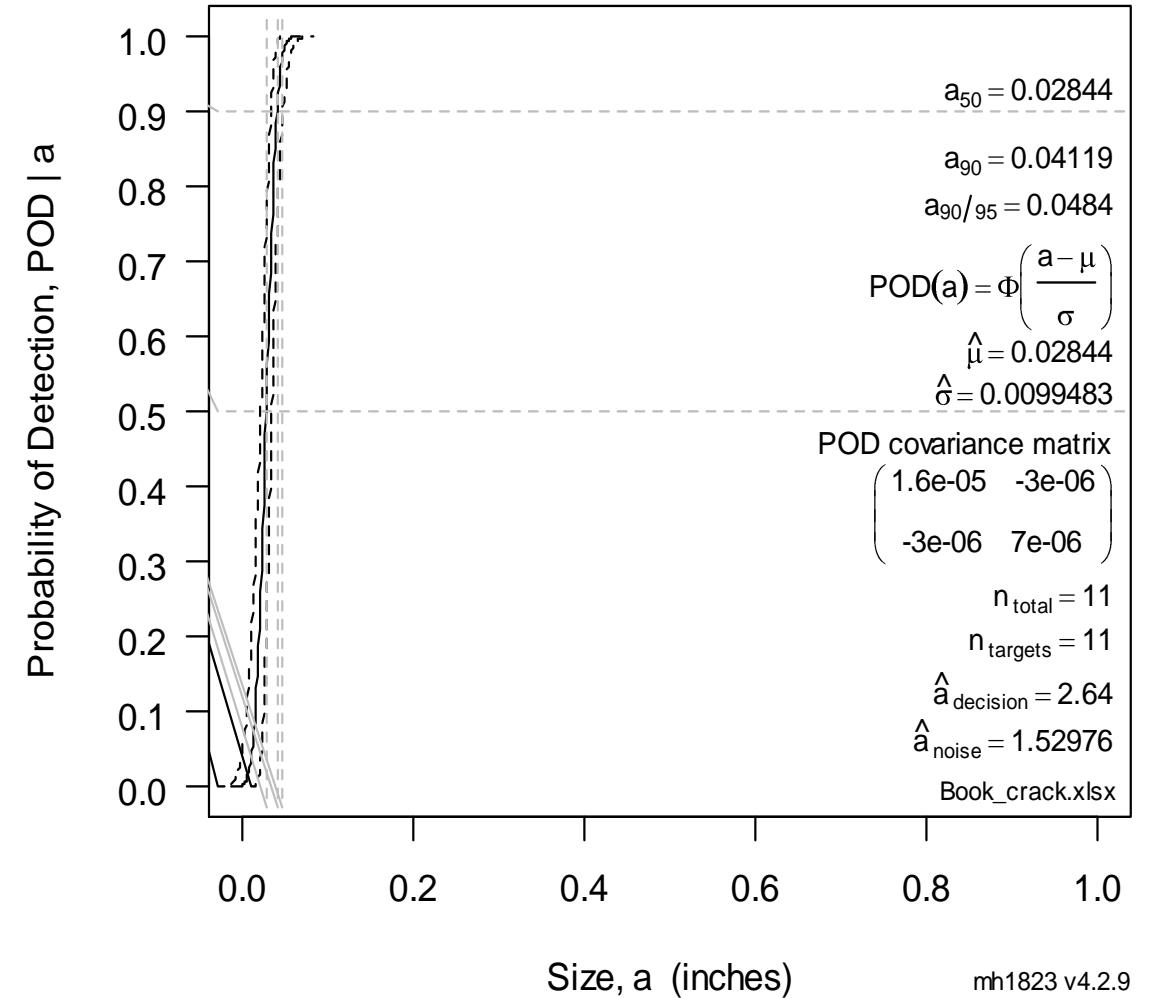
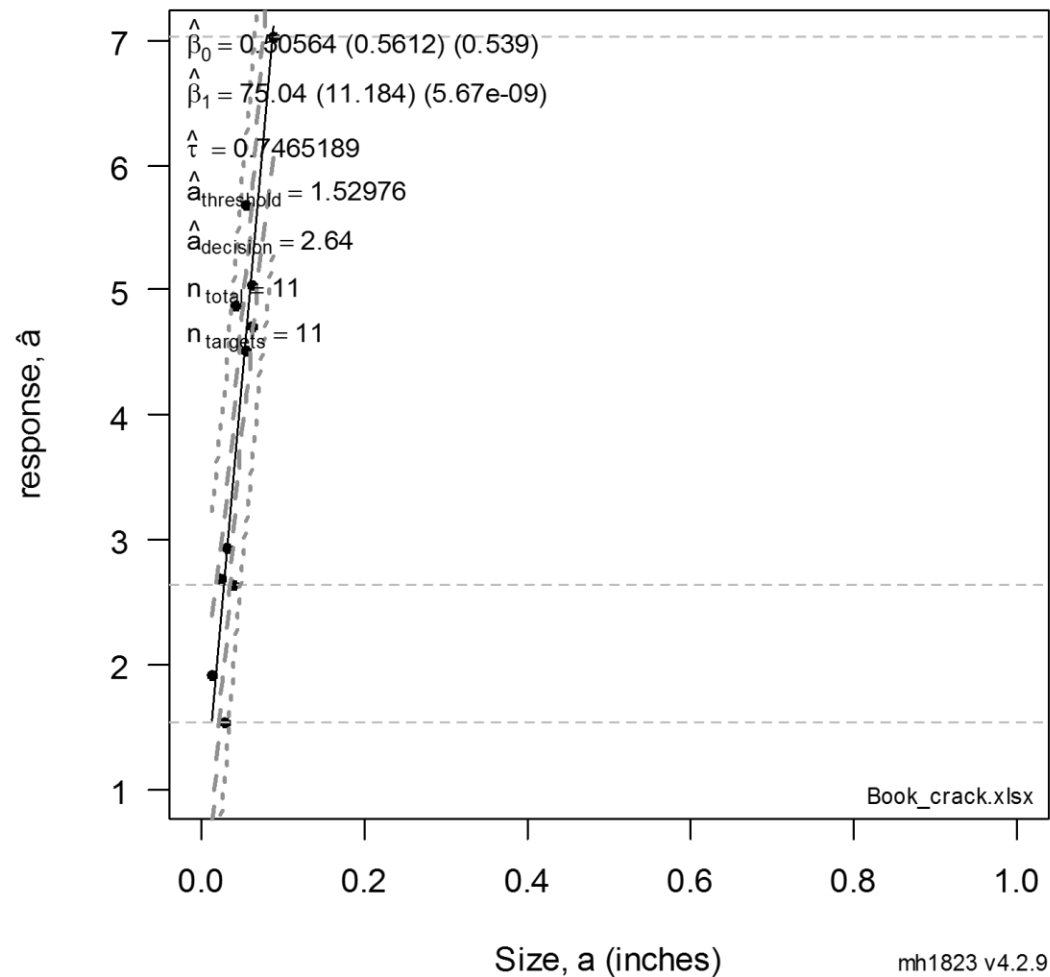
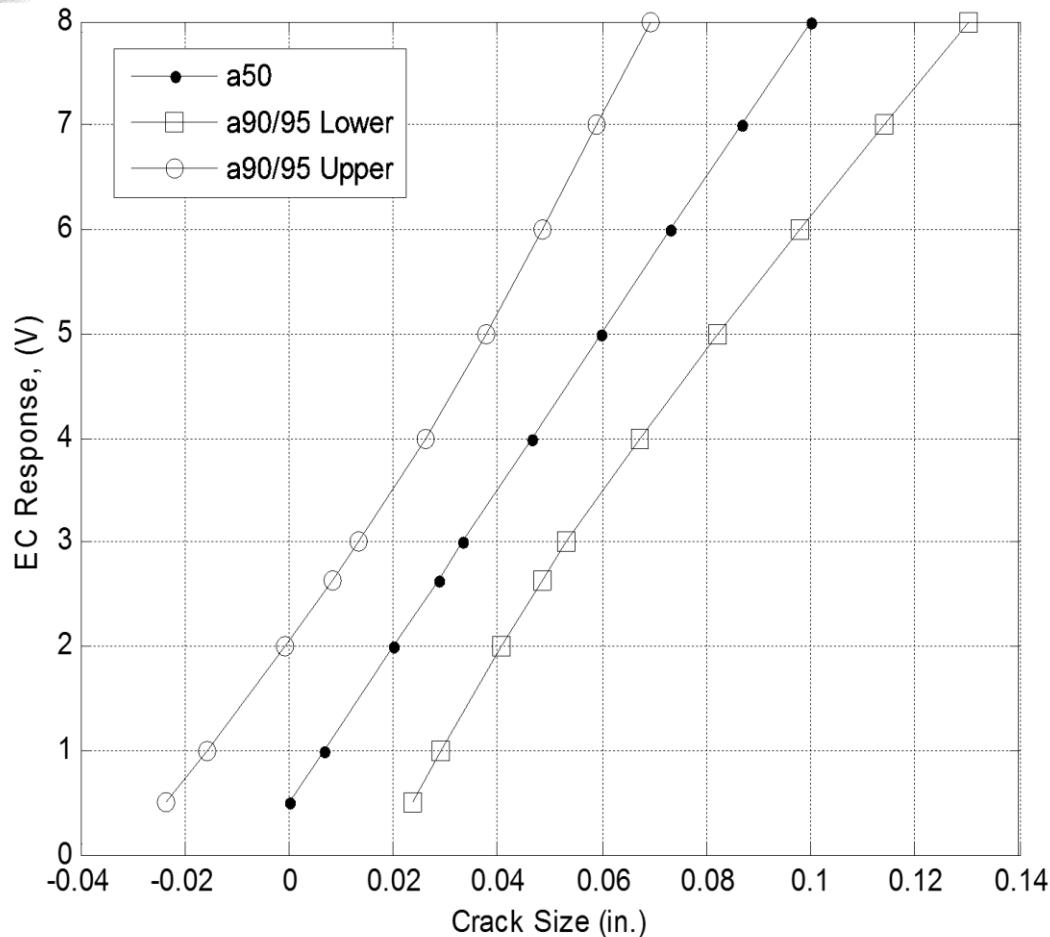
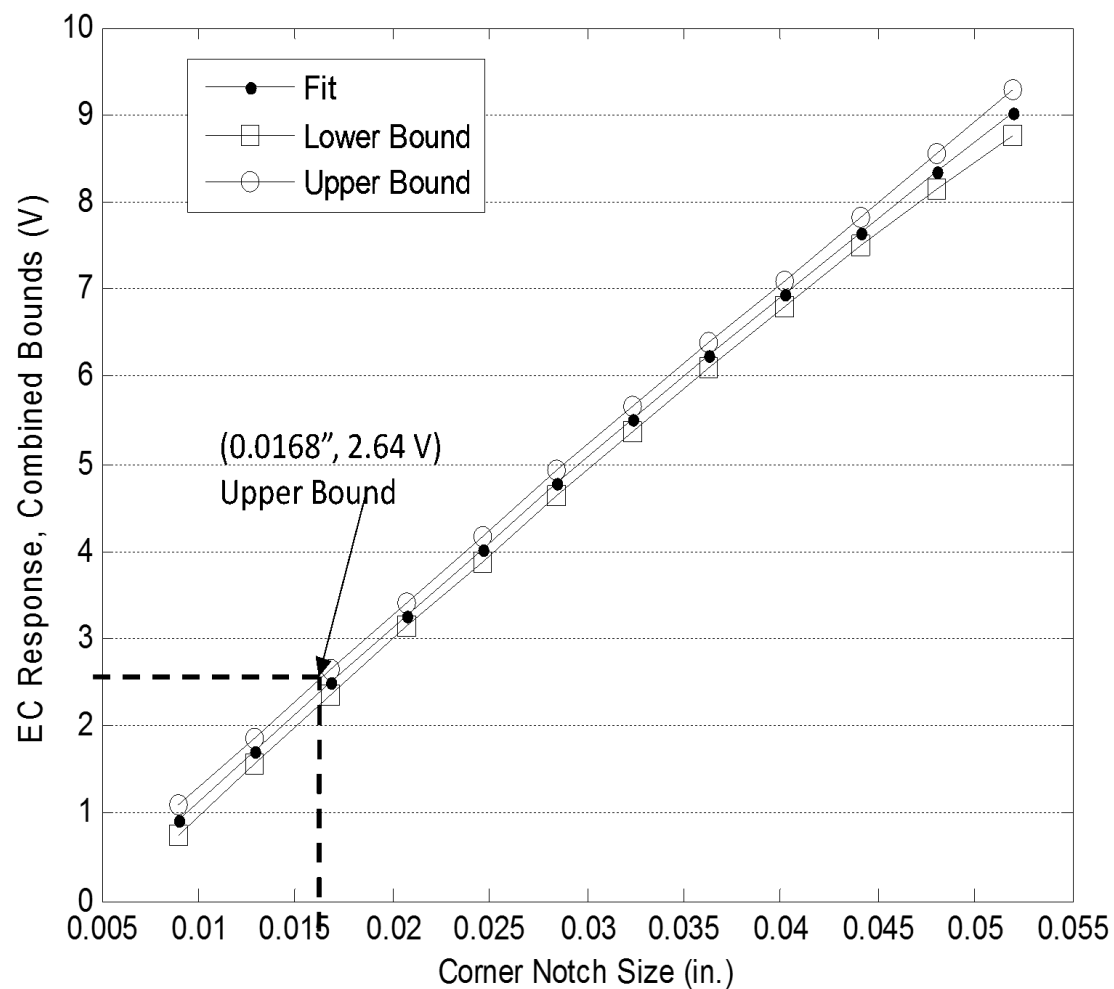


Fig. 4: mh1823 POD software calculation results per MIL-HDBK-1823.



- Alternately, we can generate several lower and upper  $a_{90/95}$  and  $a_{50}$  estimates corresponding to several values of decision threshold using mh1823 POD software and calculate lower and upper  $a_{90/95}$ .
- These  $\hat{a}$  versus  $a_{90/95}$  and  $a_{50}$  curves are plotted in Fig. 5.

Fig. 5: EC response versus lower and upper  $a_{90/95}$  calculated values using mh1823 POD software



- Case 2 builds on Case 1 by adding calibration notch response dataset. This dataset was also obtained experimentally.
- It is assumed that parts to be inspected are similar to flat plates for NDE purposes.
- Here, again we establish the combined upper and lower  $a_{90/95}$  or  $a_{90/95approx}$  bounds.

Fig. 6: Flat plate notch response fit and combined bounds

### Method A: Calculate reliably detectable flaw size for selected decision threshold

Step 1: Select decision threshold using notch standard calibration notch e.g. 2.64 V.

Step 2: Calculate lower 90/95 crack size in flat plate for 2.64 V response i.e. 0.052". See Fig. 3.

- Equivalent calibration flaw size for decision threshold of 2.64 V = 0.0168" using upper 90/95 curve for notch response. See Fig. 6.
- Ratio of reliably detectable crack size to calibration flaw size,  $R_3 = 0.052/0.0168 = 3.05$ .
- Calibration notch response for 0.052" notch = 9.032 V. See Fig. 6. Use fit value here.
- % reduction (or knockdown) on calibration response to obtain decision threshold

$$R_1 = 100 (\hat{a}_{a,notch} - \hat{a}_{decision}) / \hat{a}_{a,notch} = 100 \times (9.032 - 2.64) / 9.032 = 71\%$$

Ratio of responses is  $R_2 = \hat{a}_{decision} / \hat{a}_{a,notch} = 0.29$  or 29%.

## Method A: Calculate reliably detectable flaw size for selected decision threshold, Reading plots

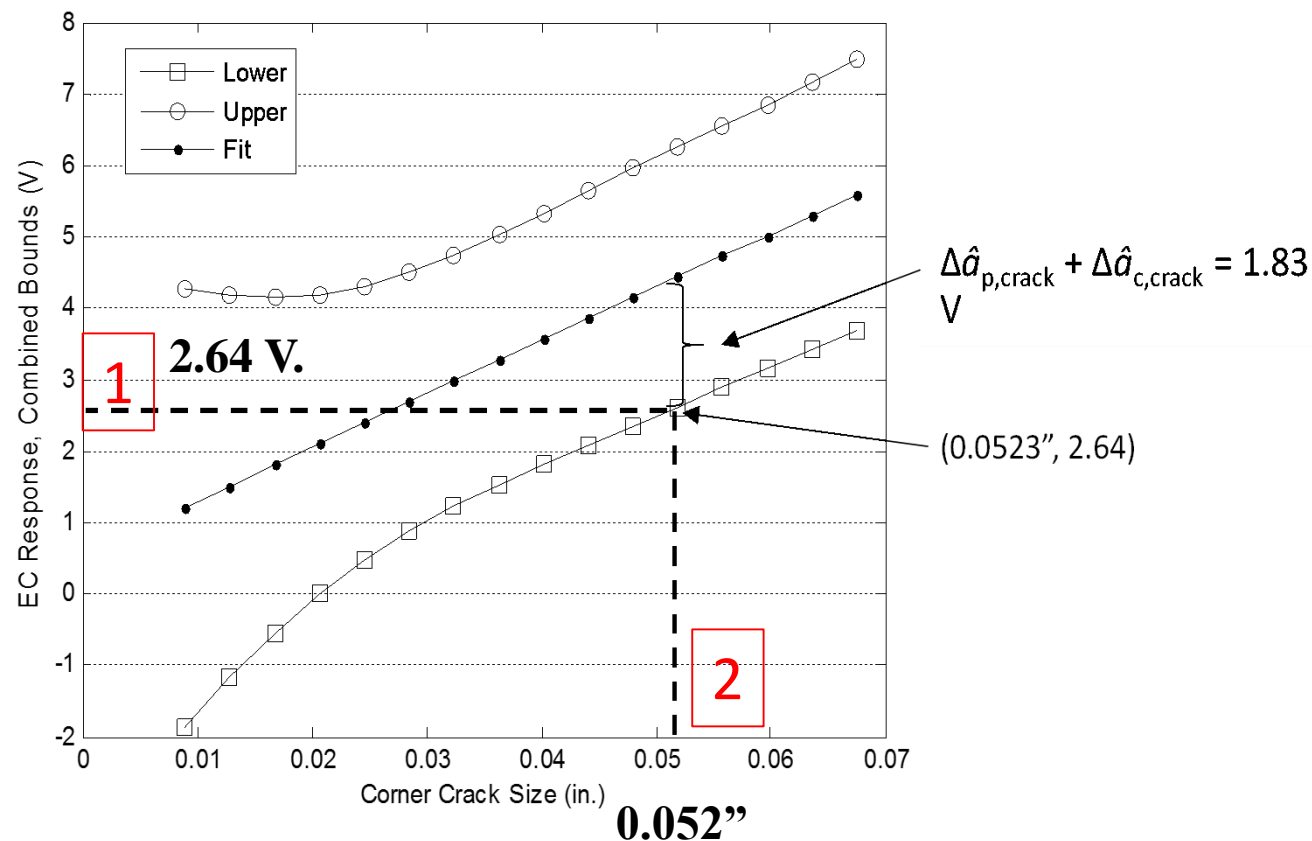
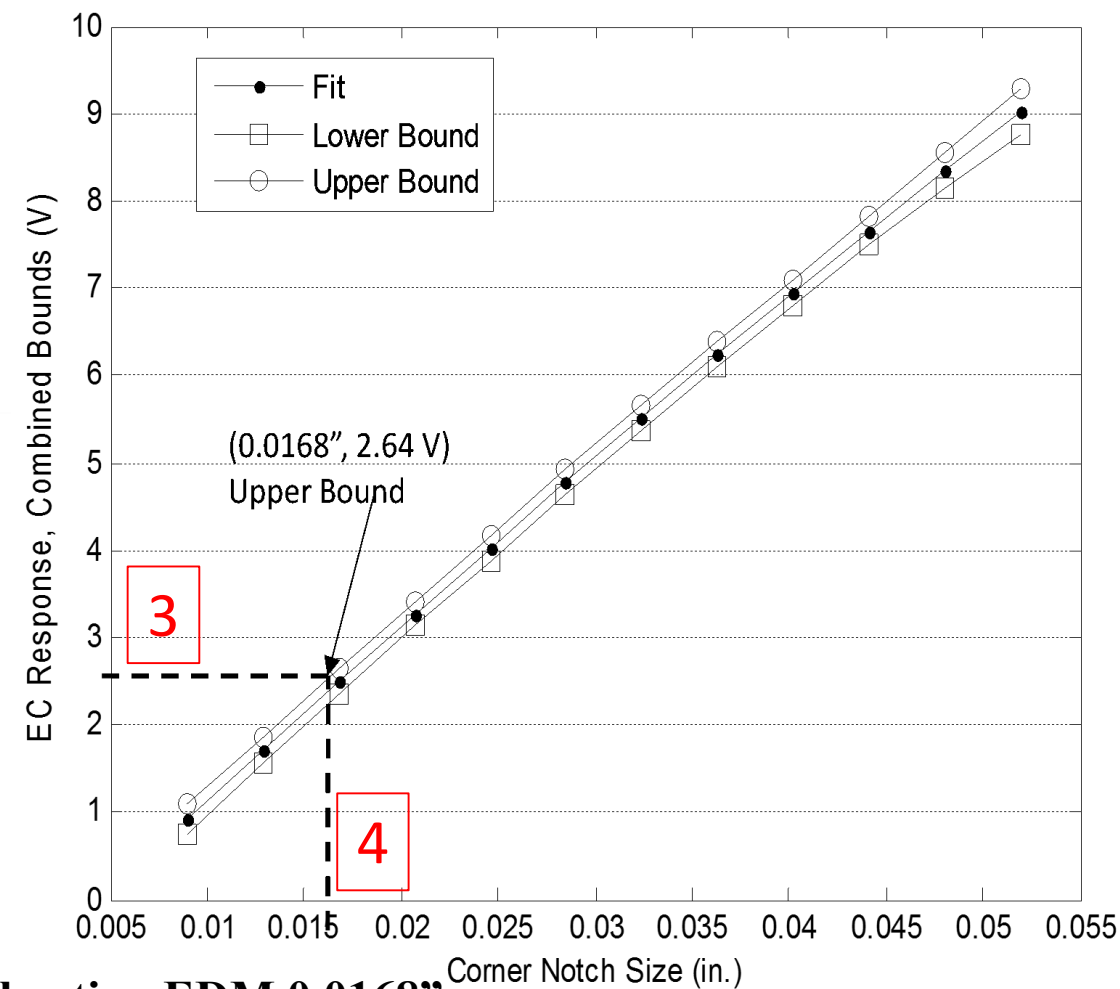


Fig. 3: Flat plate crack response fit with combined bounds



Calibration EDM 0.0168"

Fig. 6: Flat plate notch response fit and combined bounds



## Case 2: Crack and Notch Response Data Sets on Flat Plates



**Method B: Calculate decision threshold for desired reliably detectable crack size.**

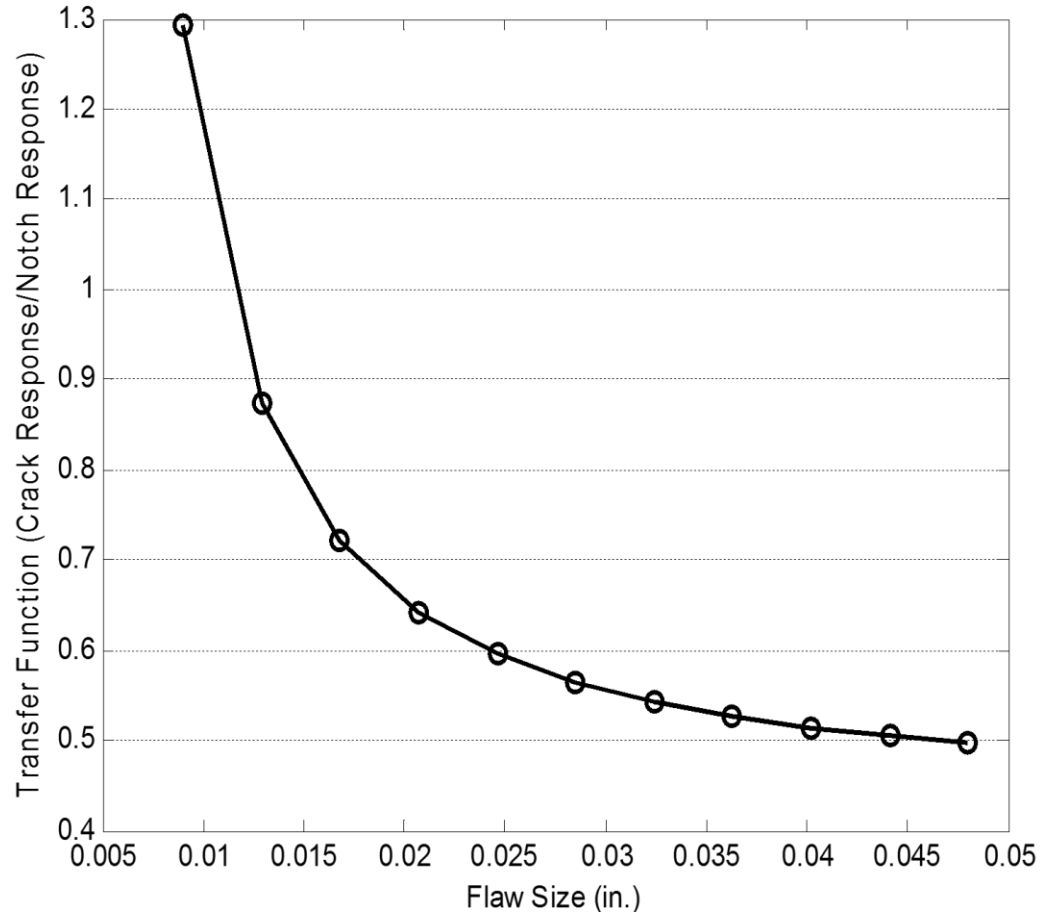


Fig. 7: Flat plate transfer function

- Before proceeding with method B, we would compute transfer function between crack response and notch response.
- Calculate crack to notch response ratio or transfer function  $R_a$  (%) by using responses from notches and cracks from flat plates. It is plotted as a function of flaw size. Transfer function is defined by following equation.

$$R_a = 100 \hat{a}_{a,crack} / \hat{a}_{a,notch}, \quad (10)$$

where,  $\hat{a}_{a,crack}$  = mean crack response for flaw size “a”, and  $\hat{a}_{a,notch}$  = mean notch response for flaw size “a”.



## Method B: Calculate decision threshold for desired reliably detectable crack size.

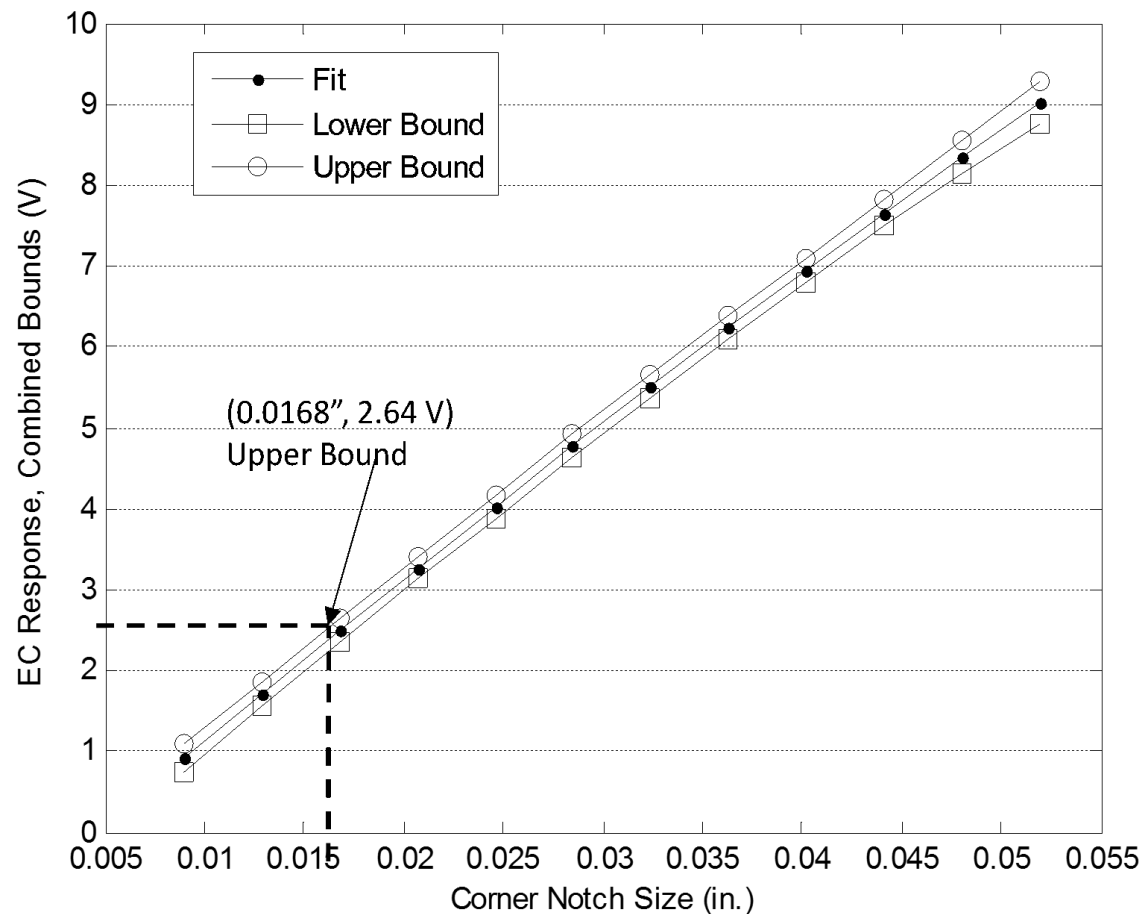


Fig. 8: EC Response from notches in flat plate

Step 1: Select crack size to be detected, i.e.  $a = 0.052''$ . See Fig. 8.

Step 2: For crack size of  $0.052''$ , calculate  $\hat{a}_{90/95 \text{ notch}} = 8.766 \text{ V}$ . See Fig. 8.

Step 3: For crack size of  $0.052''$ , calculate  $R_a = 49\%$ . See Fig. 7.

Step 4. For crack size of  $0.052''$ , Calculate  $(\hat{a}_{50 \text{ crack}} - \hat{a}_{90/95 \text{ crack}})$  or  $(\Delta\hat{a}_{p \text{ crack}} + \Delta\hat{a}_{c \text{ crack}}) = 1.83 \text{ V}$ . See Fig. 3,

where,  $\Delta\hat{a}_{p \text{ crack}}^a$  = half prediction bound interval for crack size “a”, and  $\Delta\hat{a}_{c \text{ crack}}^a$  = half confidence bound interval for crack size “a”.

Calculate any of the following two expressions.

$(\hat{a}_{90/95 \text{ notch}} R_a) / 100 = 4.295 \text{ V}$  or  
 $(\hat{a}_{a \text{ notch}} - \Delta\hat{a}_{p \text{ notch}} - \Delta\hat{a}_{c \text{ notch}}) R_a / 100 = 4.295 \text{ V}$ ,  
 where,  $\hat{a}_{a \text{ notch}}$  = notch response for notch size “a”,  
 $\Delta\hat{a}_{p \text{ notch}}$  = half prediction bound interval for notch size “a”, and  
 $\Delta\hat{a}_{c \text{ notch}}$  = half confidence bound interval for notch size “a”.

Use any of following two equations and calculate  $\hat{a}_{\text{decision}}$

$$\hat{a}_{\text{decision}} = (\hat{a}_{90/95 \text{ notch}} R_a) / 100 - (\hat{a}_{50} - \Delta\hat{a}_{90/95}), \text{ or} \quad (11)$$

$$\hat{a}_{\text{decision}} = (\hat{a}_{a \text{ notch}} - \Delta\hat{a}_{p \text{ notch}} - \Delta\hat{a}_{c \text{ notch}})(R_a / 100) - \Delta\hat{a}_{p \text{ crack}} - \Delta\hat{a}_{c \text{ crack}} \quad (12)$$

$$\hat{a}_{\text{decision}} = 4.295 - 1.77 = 2.525 \text{ V}.$$

**Method B: Calculate decision threshold for desired reliably detectable crack size 0.052".**

**Reading plots. 1 -> 2**

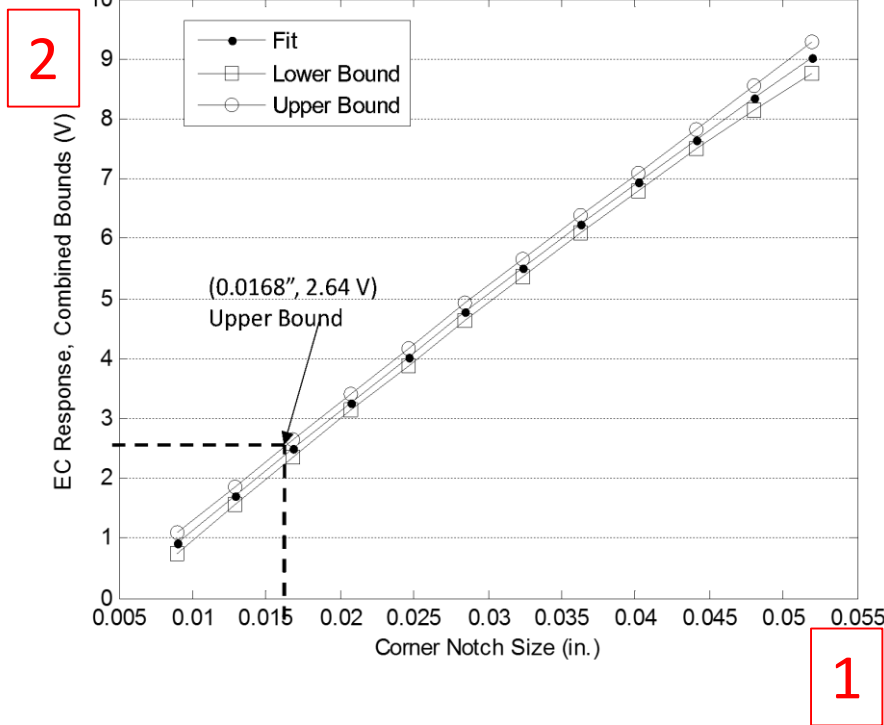


Fig. 8: EC Response from notches in flat plate

$$\hat{a}_{90/95 \text{ notch}} = 8.766 \text{ V}$$

$$R_a = 49\%.$$

$$\hat{a}_{\text{decision}} = (\hat{a}_{90/95 \text{ notch}})(R_a/100) - (\Delta \hat{a}_{p, \text{crack}} + \Delta \hat{a}_{c, \text{crack}})$$

$$\hat{a}_{\text{decision}} = 4.295 - 1.77 = 2.525 \text{ V.}$$

**3 -> 4**

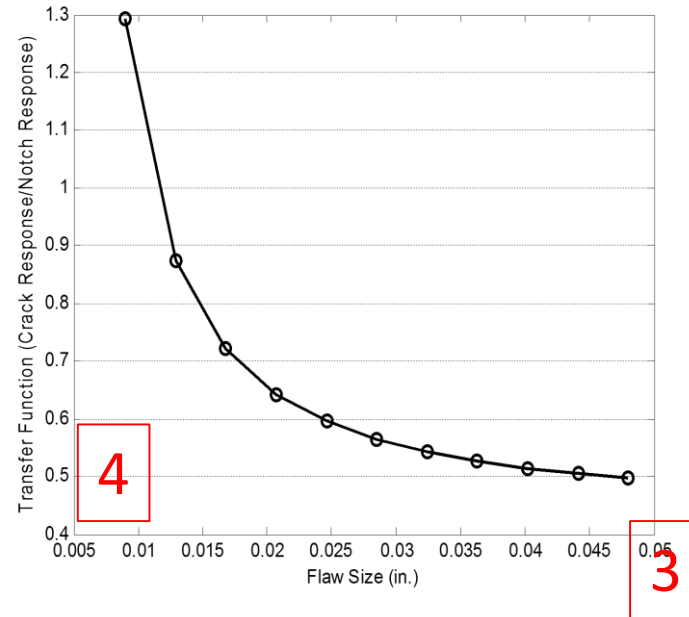


Fig. 7: Flat plate transfer function

**5 -> 6**

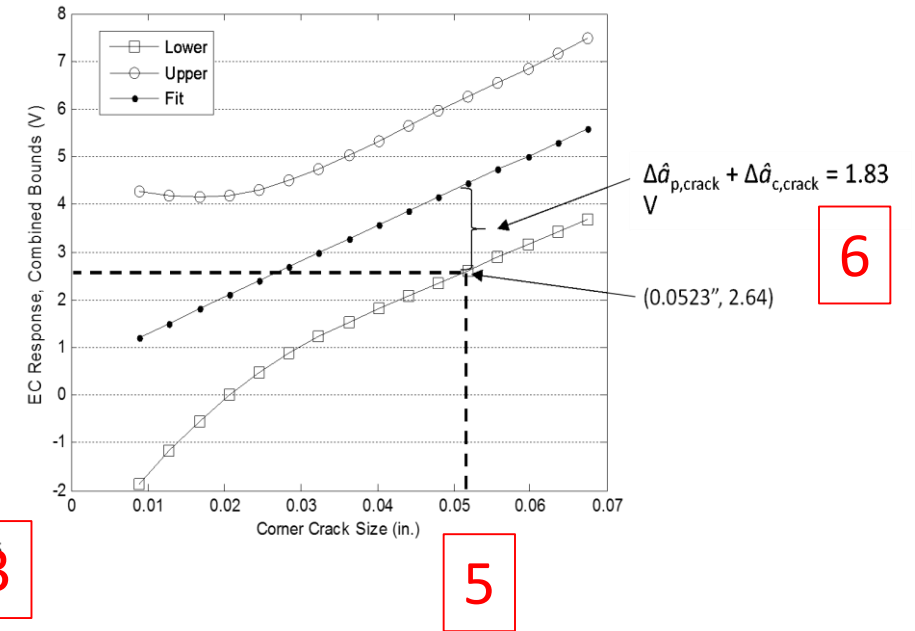


Fig. 3: Flat plate crack response fit with combined bounds

$$(\Delta \hat{a}_{p, \text{crack}} + \Delta \hat{a}_{c, \text{crack}}) = 1.83 \text{ V}$$

(12)

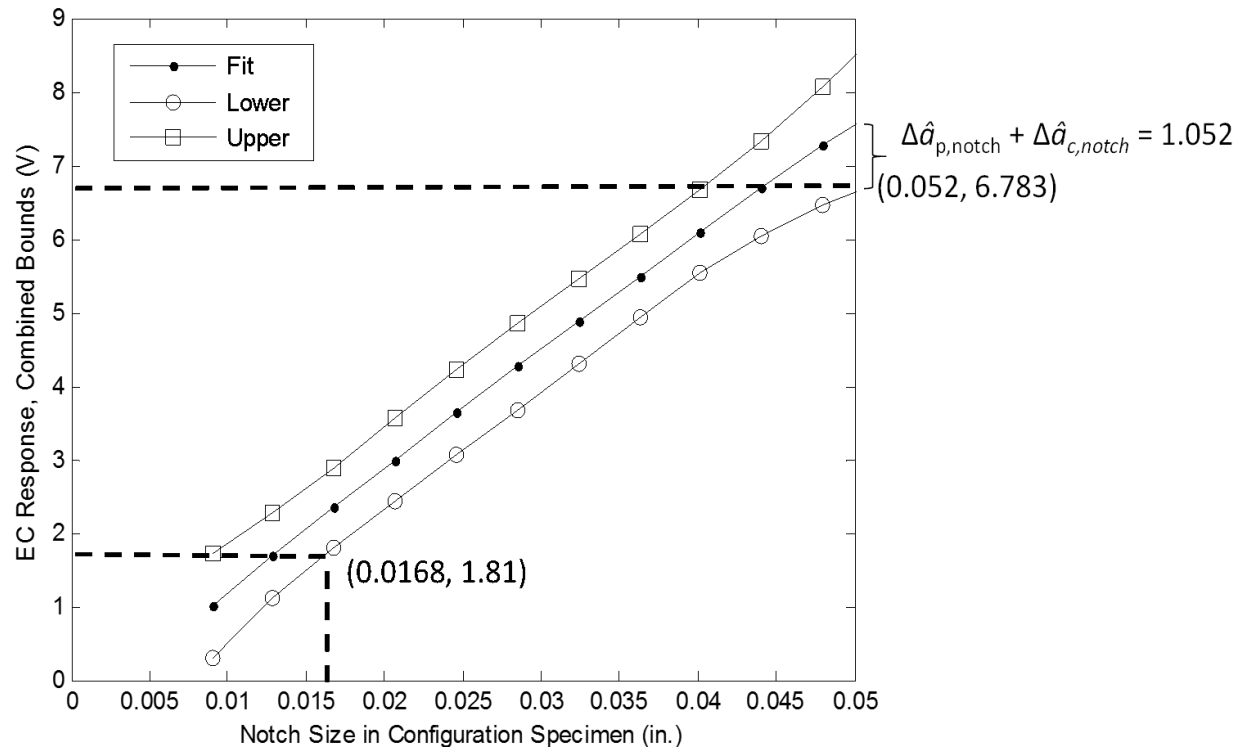
### Method B: Calculate decision threshold for desired reliably detectable crack size.

- Notch size ( $a_{e\,notch}$ ) with response equivalent to decision threshold 2.525 V is 0.0168”.
- See Fig. 8. Use fitted curve.
- % reduction or knockdown on calibration response to obtain decision threshold is given by,
- $R_1 = 100 (\hat{a}_{a,notch} - \hat{a}_{decision}) / \hat{a}_{a,notch} = 100 \times (9.032 - 2.525) / 9.032 = 72\%$ .
- Ratio of decision threshold to calibration notch response,  $R_2 = \hat{a}_{decision} / \hat{a}_{a,notch} = 0.28$  or 28 %.
- Ratio of detection flaw size to equivalent calibration flaw size,  $R_3 = a / a_{e\,notch} = 0.052'' / 0.0168'' = 3.1$ .

**Table 1: Method A and B calculation results for reliable detection of 0.052” crack for Case 2**

Method	Decision Threshold, V	Reduction in Calibration Flaw Response to Obtain Decision Threshold $R_1$ , %	Ratio of Decision Response to Calibration Response, $R_2$	Ratio of Detection Flaw Size to Equivalent Calibration Flaw Size, $R_3$
A	2.6	71	29%	3.05
B	2.5	72	28%	3.1

## Configuration specimen notch response with 90/95 bounds



- Consider eddy current surface inspection of welded tubing.
- Here, we assume that either it is not practical to make known size crack specimens in welded tubing or it is too costly.
- Only practical solution is to make artificial flaws such as EDM notches in specimens with part configuration.

Fig. 9: Configuration specimen notch response with 90/95 bounds

## Comparison of configuration specimen notch response with flat plate notch response

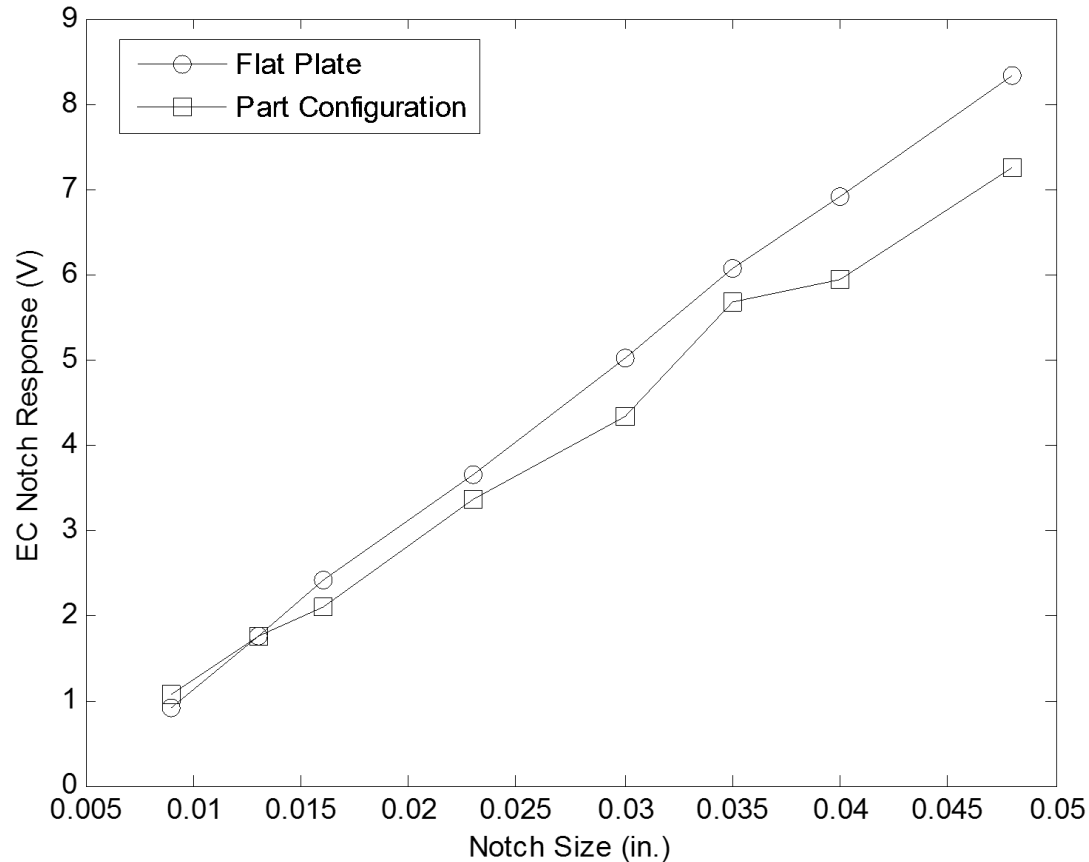


Fig. 10: Comparison of configuration specimen notch response with flat plate notch response

- The comparison shall indicate that same type of GLM or other model equation would work.
- Mean difference between the two curves should be constant or monotonic with flaw size.
- Two methods are provided for estimation of reliably detectable flaw size.



## Case 3: Detection of Cracks in Complex Geometry Parts

### Method A: Determine reliably detectable crack size for a chosen calibration threshold

Step 1. Select decision threshold using configuration standard calibration notch e.g. 1.81 V.

Step 2: Calculate lower 90/95 notch size in configuration standard e.g. 0.0168". See Fig. 9.

Step 3: Calculate upper 90/95 flat plate response for 0.0168" notch i.e. 2.64 V. See Fig. 6.

Step 4: Calculate lower 90/95 crack size in flat plate for 2.64 V response i.e. 0.0523". See Fig. 3.

- Equivalent calibration response for 0.052" crack = 7.834 V. Use fit curve value. See Fig. 9.
- Decision threshold 1.81 V equivalent calibration notch size = 0.0136". See Fig. 9. Use fit curve.
- Ratio of to reliably detectable crack size to calibration flaw size,  $R_3 = 0.0523"/0.0136" = 3.84$ .
- % reduction (knockdown) on equivalent calibration response to obtain the decision threshold,
- $R_1 = 100 (\hat{a}_{a,notch} - \hat{a}_{decision}) / \hat{a}_{a,notch} = 100(7.834 - 1.81)/7.834 = 77\%$ .
- Ratio of decision threshold to calibration notch response,  $R_2 = \hat{a}_{decision} / \hat{a}_{a,notch} = 0.23$  or 23 %.

# Case 3: Detection of Cracks in Complex Geometry Parts

**Method A: Determine reliably detectable crack size for a chosen calibration threshold 1.81 V.**

1 -> 2 -> 3 -> 4 -> 5 -> 6    7->8, 9 ->10

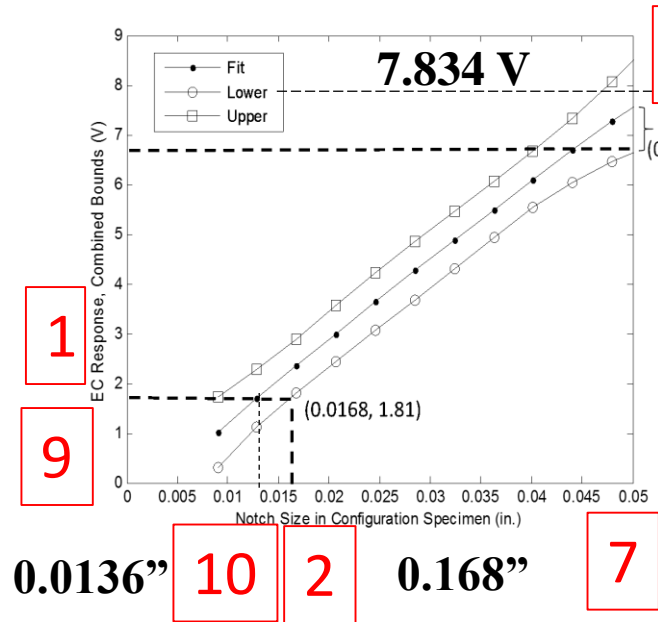


Fig. 9: Configuration specimen notch response with 90/95 bounds

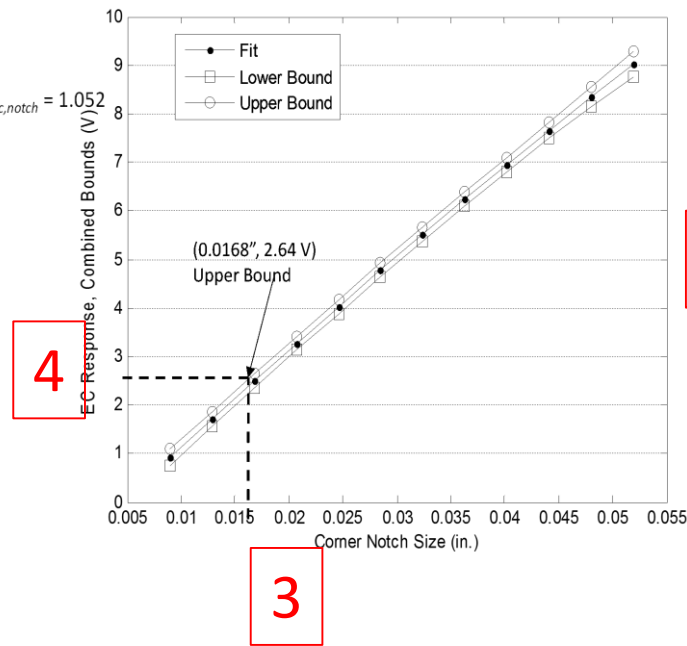


Fig. 6: Flat plate notch response fit and combined bounds

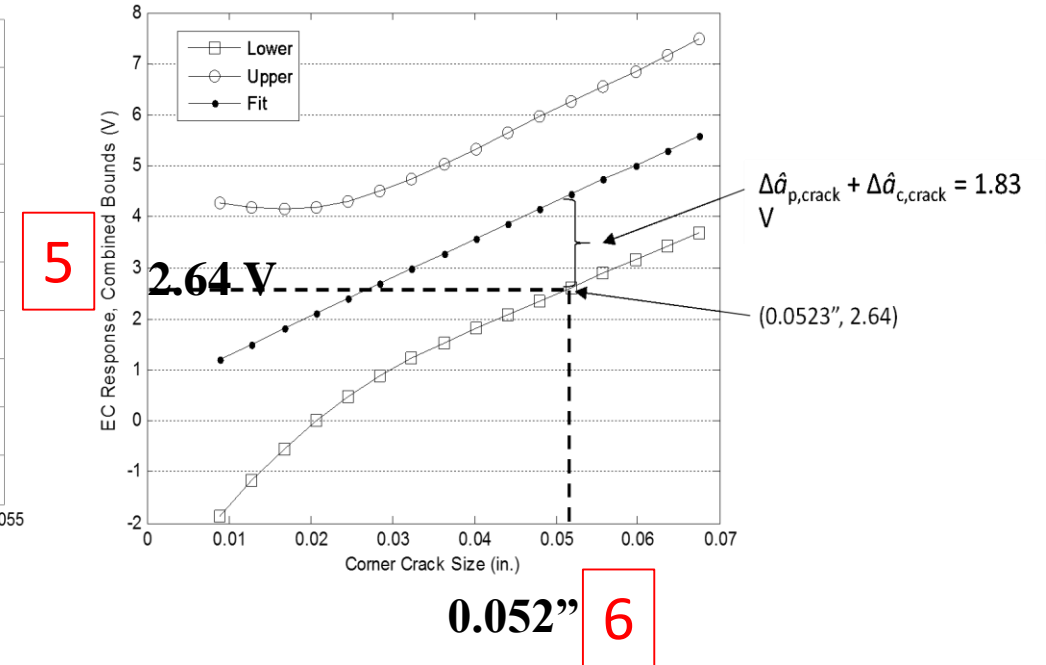


Fig. 3: Flat plate crack response fit with combined bounds





## Case 3: Detection of Cracks in Complex Geometry Parts



### Method B: Calculate decision threshold for desired reliably detectable crack size.

Method B uses transfer function. Method B assumes that variance in flat plate notch data is negligible.

Objective: Calculate decision threshold for desired reliably detectable crack size.

Step 1. Select crack size to be detected,  $a = 0.052''$ . See Fig. 9.

Step 2. Calculate configuration specimen EDM notch lower  $\hat{a}_{90/95}$  for the selected crack size.  $\hat{a}_{90/95} = \hat{a}_{a,notch} - \Delta\hat{a}_{p,notch} - \Delta\hat{a}_{c,notch} = 6.783V$ . See Fig. 9.

Step 3. Calculate  $R_a = 49\%$  for  $0.052''$  crack. See Fig. 7.

Step 4. Calculate  $(\hat{a}_{90/95,crack} - \hat{a}_{90/95,crack})$  or  $(\Delta\hat{a}_{p,crack} + \Delta\hat{a}_{c,crack}) = 1.83V$ . See Fig. 3.

- $(\hat{a}_{90/95,a,notch} R_a) / 100 = 3.324 V$  or
- Calculate  $(\hat{a}_{a,notch} - \Delta\hat{a}_{p,notch} - \Delta\hat{a}_{c,notch}) R_a / 100 = 3.324 V$ .
- Use any of following equations and calculate,
  - $\hat{a}_{decision} = (\hat{a}_{90/95,a,notch} R_a) / 100 - (\hat{a}_{50} - \Delta\hat{a}_{90/95})$ , or
  - $\hat{a}_{decision} = (\hat{a}_{a,notch} - \Delta\hat{a}_{p,notch} - \Delta\hat{a}_{c,notch})(R_a/100) - \Delta\hat{a}_{p,crack} - \Delta\hat{a}_{c,crack}$ .
  - $\hat{a}_{decision} = 3.324 - 1.83 = 1.494V$ .
- Note: that quantities  $\hat{a}_{a,notch}$ ,  $\Delta\hat{a}_{p,notch}$  and  $\Delta\hat{a}_{c,notch}$  are for notches in configuration specimens. % reduction (knockdown) on calibration response to obtain the decision threshold is calculated as follows,
  - $R_1 = 100 (\hat{a}_{a,notch} - \hat{a}_{decision}) / \hat{a}_{a,notch} = 100 \times (7.835 - 1.494) / 7.835 = 80.9\%$ .
  - Ratio of decision threshold to calibration notch response  $R_2 = \hat{a}_{decision} / \hat{a}_{a,notch} = 0.19$  or  $19\%$ .
- Equivalent calibration flaw size that provides decision threshold response =  $0.012''$ . See Fig. 9. Use fitted curve.
- Ratio of detection crack size to equivalent calibration flaw size  $R_3 = 0.052'' / 0.012'' = 4.3$

## Method B, Calculate decision threshold for desired reliably detectable crack size 0.052"

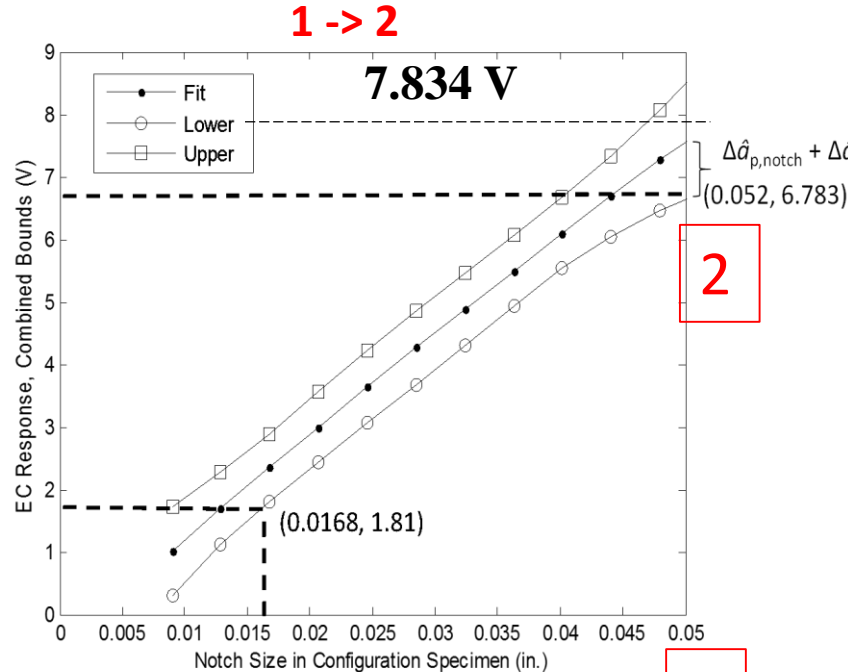


Fig. 9: Configuration specimen notch response with 90/95 bounds

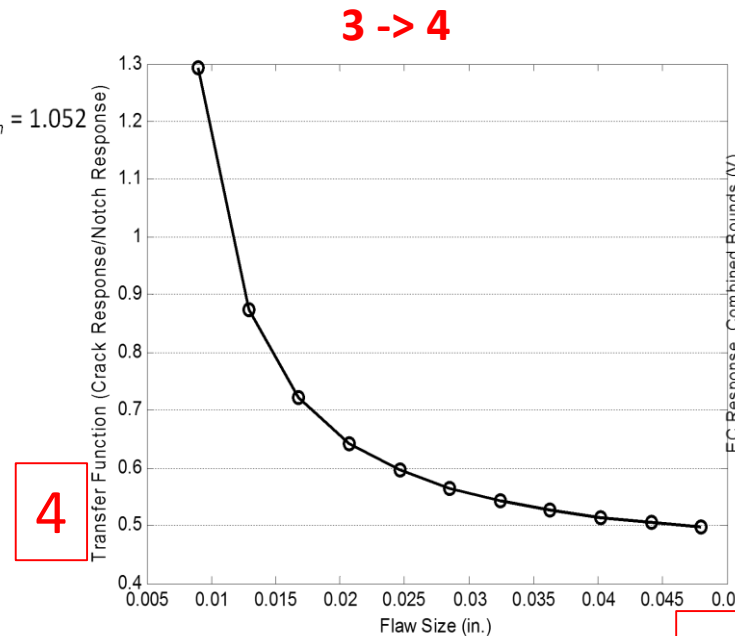


Fig. 7: Flat plate transfer function

$$R_a = 49\%$$

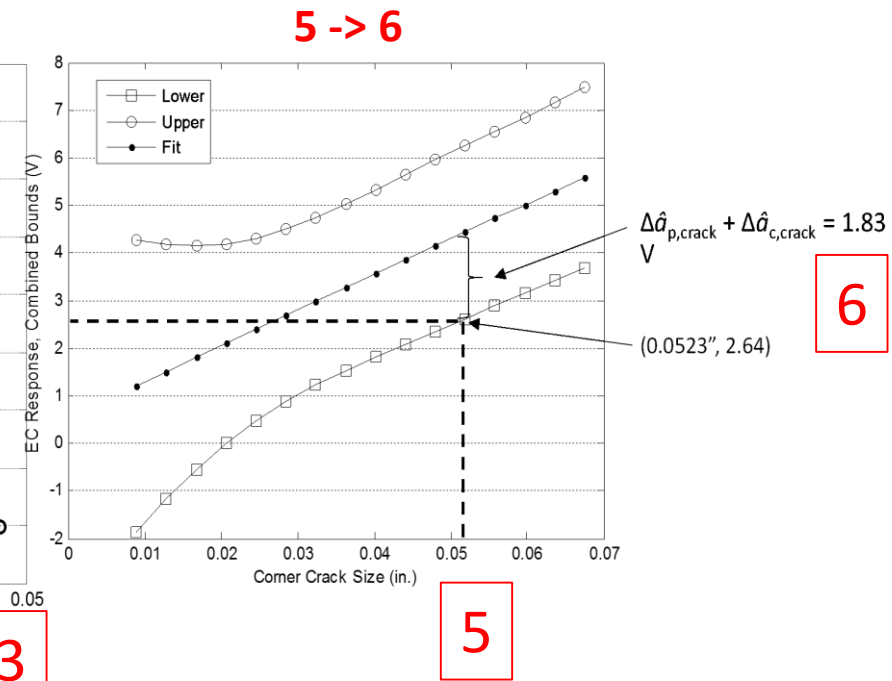


Fig. 3: Flat plate crack response fit with combined bounds

$$(\Delta \hat{a}_{p,crack} + \Delta \hat{a}_{c,crack}) = 1.83 \text{ V}$$

$$\hat{a}_{\text{decision}} = (\hat{a}_{a,notch} - \Delta \hat{a}_{p,notch} - \Delta \hat{a}_{c,notch})(R_a/100) - (\Delta \hat{a}_{p,crack} + \Delta \hat{a}_{c,crack}) = 1.494.$$

Note: that quantities  $\hat{a}_{a,notch}$ ,  $\Delta \hat{a}_{p,notch}$  and  $\Delta \hat{a}_{c,notch}$  are for notches in configuration specimens.



## Case 3: Detection of Cracks in Complex Geometry Parts



**Table 2: Method A and B calculation results for reliable detection of 0.052” crack for Case 3**

Method	Decision Threshold, V	Reduction in Calibration Flaw Response to Obtain Decision Threshold $R_1$ , %	Ratio of Decision Response to Calibration response, $R_2$	Ratio of Detection Flaw Size to Equivalent Calibration Flaw Size, $R_3$
A	1.81	77	23%	3.84
B	1.49	81	19%	4.3



# Noise Analysis for Probability of False Calls

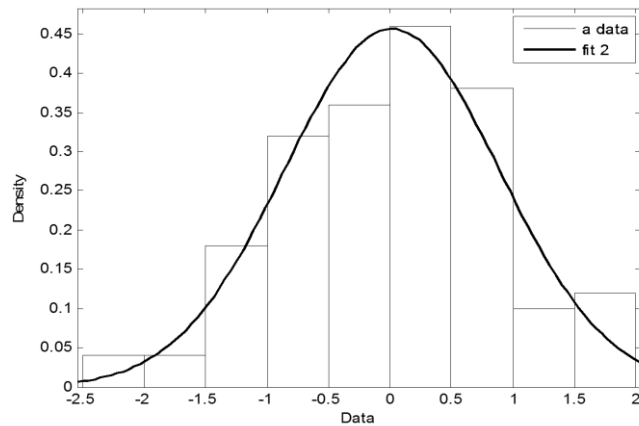


Fig. 11: Distribution function fit to noise data

Fit: Maximum Likelihood Estimation (MLE),  
Distribution: Normal,  
Log likelihood: -143.012,  
Domain:  $-\infty < y < \infty$ ,  
Mean,  $\mu$ : -0.110793,  
Variance: 1.0329,  
Sigma: 1.01631.

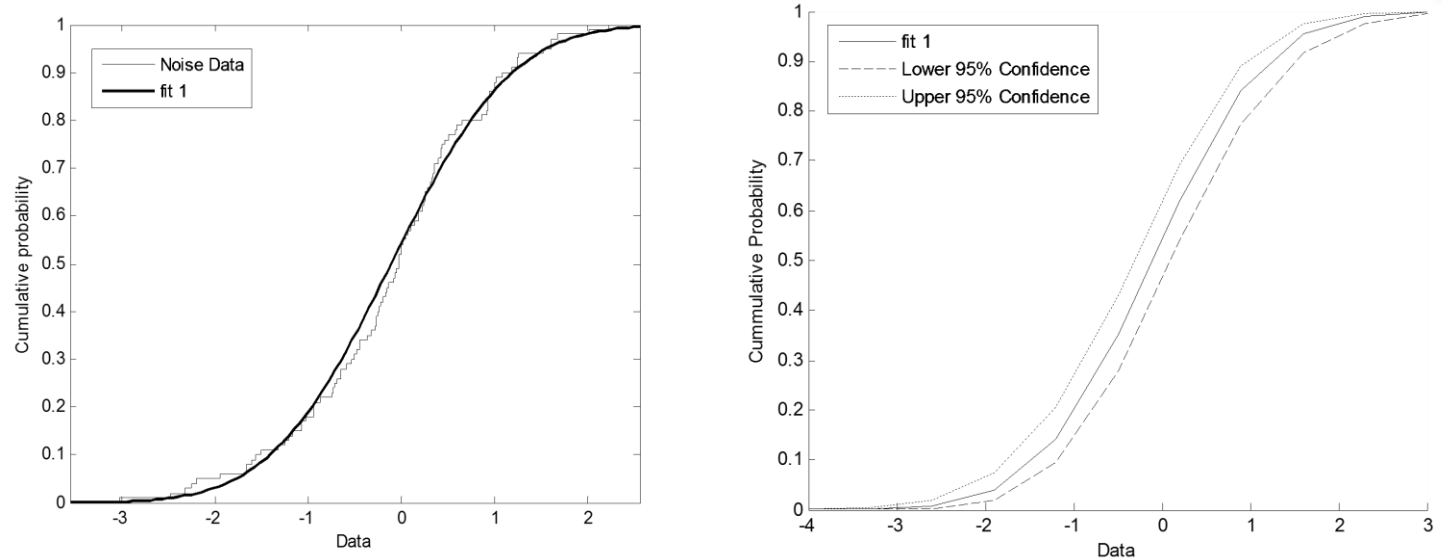
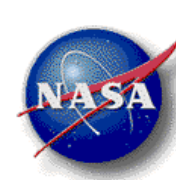


Fig. 12: a) Cumulative noise data and cumulative probability density function, b) Cumulative probability density function or  $POD_{noise}$  and confidence bounds.

- Probability of false calls (POF) is given by  $POD_{noise}$ , where noise level is equal to the decision threshold.

$$POF = 1 - POD_{noise} \quad (13)$$

- Refer to MIL-HDBK-1823 for in depth analysis of noise and estimation of probability of false calls.



# Conclusions



- Paper provides analysis methods and procedures to estimate reliably detectable flaw size for chosen decision threshold or to estimate decision threshold associated with given reliably detectable flaw size for NDE methods that use calibration on artificial flaws.
- The approach assumes positive and high correlation (i.e.  $R\text{-square} > 0.8$ ) of signal response with flaw size for all datasets used in the analysis.
- The procedures use  $a_{90/95}$  or  $a_{90/95approx}$  bounds of input datasets in the analysis.
- Accordingly, the resulting reliably detectable flaw sizes also have 90% POD with high confidence that can be approximately equal to 95% based on validity of assumptions.
- To complete the POD analysis, paper also provides information on estimating of probability of false calls.
- Reduction in calibration flaw response to obtain decision threshold  $R_1$ , % was over 70% in the example calculations for 90/95 POD/Confidence.